Developmental Cognitive Neuroscience 33 (2018) 149–160

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn

Methodological considerations for developmental longitudinal fMRI research

Eva H. Telzer[a,*], Ethan M. McCormick[a,1], Sabine Peters[b,c,1], Danielle Cosme[d], Jennifer H. Pfeifer[d], Anna C.K. van Duijvenvoorde[b,c]

[a] University of North Carolina, Chapel Hill, USA
[b] Leiden University, The Netherlands
[c] Institute of Psychology, Leiden University, Leiden, The Netherlands
[d] University of Oregon, Eugene, USA

ARTICLE INFO

Keywords:
Longitudinal fMRI
Development
Methods

ABSTRACT

There has been a large spike in longitudinal fMRI studies in recent years, and so it is essential that researchers carefully assess the limitations and challenges afforded by longitudinal designs. In this article, we provide an overview of important considerations for longitudinal fMRI research in developmental samples, including task design, sampling strategies, and group-level analyses. We first discuss considerations for task designs, weighing the pros and cons of many commonly used tasks, as well as outlining how the tasks may be impacted by repeated exposure. Secondly, we review the types of group-level analyses that can be conducted on longitudinal fMRI data, analyses which must account for repeated measures. Finally, we review and critique recent longitudinal studies that have emerged in the past few years.

1. Introduction

Functional magnetic resonance imaging (fMRI) research in developmental populations has spiked in the past 20 years, with the majority of published work coming out in the past 5 years (Herting et al., this issue). This research has significantly advanced our understanding of the neural processes that support social, cognitive, and affective changes from childhood to adulthood (Crone and Elzinga, 2015). Most developmental fMRI research to date has utilized cross-sectional samples, which examines differences and similarities in neural activation between children, adolescents, and adults. However, cross-sectional studies are limited in their ability to examine how maturation of brain function develops *within* individuals. More recently, developmental cognitive neuroscientists have moved towards implementing various longitudinal designs, which is necessary for truly unpacking developmental processes. Longitudinal fMRI offers the advantage of removing between-subject variability, instead using the individual as their own control (Louis et al., 1986), which increases our ability to separate developmental effects from cohort effects and reduces sampling biases (Crone and Elzinga, 2015). Additionally, longitudinal fMRI does not make assumptions about the stability of brain-behavior relationships and is particularly well suited to detect developmental transitions (McCormick et al., 2017). Finally, cross-sectional studies are not suitable for testing mediating effects, for instance, examining whether alterations in brain development explain links between early life stress and later psychopathology. Thus, to test causal pathways and identify neural mechanisms longitudinal studies are needed (Crone and Elzinga, 2015).

Despite its many advantages, longitudinal fMRI presents many challenges. First and most obviously, longitudinal studies are costly, time-consuming, and complicated. Researchers must carefully manage subject retention, as the validity of findings from longitudinal fMRI can be undermined by participant dropout over time. Moreover, while cross-sectional studies can compare children to adolescents and adults in a very short period of data collection, longitudinal studies require many years to capture this same window. One solution that researchers have utilized for examining developmental trajectories over a broad age range in shorter periods is an accelerated longitudinal design (also known as cohort-sequential), in which longitudinal data are collected in multiple age cohorts over time. Secondly, researchers must carefully design the fMRI tasks, as learning, practice, and habituation effects become significant confounds in longitudinal studies. Finally, longitudinal fMRI data analyses are more complicated due to the nested nature of the data, which violates the assumption of independence that

* Corresponding author.
E-mail address: ehtelzer@unc.edu (E.H. Telzer).
[1] Equal author contribution.

https://doi.org/10.1016/j.dcn.2018.02.004
Received 18 September 2017; Received in revised form 31 January 2018; Accepted 5 February 2018
Available online 07 February 2018

underlies many standard statistical packages for fMRI analyses.

In this article, we provide an overview of important considerations and limitations for longitudinal fMRI research in developmental samples, including task design, sampling strategies, and longitudinal group-level analyses. It is essential that researchers carefully assess the limitations but also opportunities afforded by longitudinal designs. Below we first discuss the considerations for task designs, weighing the pros and cons of many commonly used tasks, as well as outlining how the tasks may be impacted by repeated exposure. Secondly, we review the types of group-level analyses that can be conducted on longitudinal fMRI data, analyses which must account for repeated measures. Finally, we review and critique recent longitudinal studies that have emerged in the past few years.

2. Task considerations for longitudinal neuroimaging

Of all the decisions that need to be made in a developmental longitudinal neuroimaging study, selection of the appropriate task parameters is perhaps the most important in that it constrains every future decision that can be made concerning data processing and analysis. Neuroimaging tasks are, at their core, simply a tool designed to measure the neural representations underlying a psychological construct (e.g. inhibitory control). When adding a longitudinal component, a task must also be able to capture changes in brain and behavior that are associated with the development of the psychological process over time. We outline several considerations that should be taken into account when selecting a task for longitudinal neuroimaging studies, including simultaneous changes in brain and behavior, motivational consistency, habituation to stimuli, deception, and learning. We of course recognize that the described tasks are merely a sample of possible tasks that can be used in a longitudinal neuroimaging paradigm; however, the principles outlined should apply to a wide range of potential tasks. Ultimately it is up to the researcher to know their task-of-choice best and to account for the particulars in such a way as to best answer the research question of interest.

2.1. Basic cognitive tasks

Tasks that test participants' basic cognitive abilities (e.g., working memory, cognitive control, attention, etc.) have a long history in neuroimaging research and are often characterized by their relative simplicity and robustness in eliciting representational patterns of activation. Their application to longitudinal neuroimaging may seem prima facie a relatively straight-forward proposition, however, repeated measures can introduce unique confounds that need to be addressed when designing a task. To illustrate, we will compare two basic cognitive tasks aimed at assessing inhibitory control, the Go/Nogo (GNG; Menon et al., 2001) and the Stop Signal tasks (SST; Li et al., 2006). Both the GNG and SST measure inhibitory control by having participants respond to one type of cue, while withholding that response when presented with a second type of cue (e.g. respond to every letter but X, respond to each cue unless a tone is played, etc.). A failure to inhibit the pre-potent response when presented with the second cue is termed a false alarm. The two tasks differ primarily in that the SST adapts to participants' behavioral performance whereas task difficulty is held constant on the GNG; that is, the SST becomes harder for participants who perform well, and easier for participants who perform poorly. The result of this adaptation is that every participant should fail on half the trials, resulting in an equal number of successful inhibitions and false alarms. Inhibitory control is indexed by the difficulty level at which participants are committing 50% false alarms. In contrast, the relatively simple ratio of successful inhibitions to false alarms on the GNG reflects each participant's average inhibitory control ability. Finally, the GNG relies on a sequence of Go stimuli to develop the pre-potent response, while the SST generates this response on the trial level by presenting the signal to stop after signaling a go response at various delays.

While the metric of inhibitory control is comparatively robust and easy to interpret, the GNG task design presents several challenges when being used to assess within-person change. First, the number of successful inhibition trials is a metric of performance as well as a determinant of how much data is available to estimate the neural signature associated with successful or unsuccessful inhibition. This confound makes it difficult to determine if changes seen developmentally are the result of changes in behavior or changes in the neural systems supporting inhibitory control. If behavioral performance improves between waves (i.e. fewer false alarms), for instance, there will be relatively better estimation of successful inhibition trials and relatively poorer estimation of unsuccessful trials. Furthermore, better performance at later waves may mean that successful inhibition trials required less effortful control over time, making trials ostensibly of the same type difficult to compare across waves (i.e., greater task difficulty for younger participants versus less effortful for older participants). One approach to resolving these difficulties is to adapt the difficulty of the GNG at each wave, for instance by including 40% no-go trials at earlier waves and 30% no-go trials at later waves, ostensibly equating task difficulty across development (see Durston et al., 2002 for cross-sectional example). However, this is a relatively coarse approach and does not take into account individual differences in the development of inhibitory control. The SST, alternatively, standardizes both the number of successful/unsuccessful inhibition trials and the difficulty of the task for each participant, accounting for differences between individuals within a wave, as well as between the same person across repeated measures. This is especially helpful in avoiding floor or ceiling effects potentially caused by large swings in performance across development. Furthermore, standardization allows for consistent estimation of neural effects that reflect the same level of inhibitory effort within an individual across time. Finally, because the SST un-yokes the change in behavior from the change in neural signal estimation, this allows one to make more robust claims about the development of neural systems involved in inhibitory control and their contribution to behavior.

Basic cognitive tasks lend themselves the most to implementing adaptive scaling similar to the SST. While not always possible, especially for tasks that fall into the proceeding sections, implementing adaptive tasks can be a powerful tool. De-coupling simultaneous neural and behavioral changes can remove barriers to interpretation and provide a more-reliable estimation of how underlying neural networks, and the cognitive skills they support, develop across time. Yet, an important consideration for using adaptive tasks is behavioral performance can become more challenging to measure. For instance, if a task is adaptive such that one performs at a 50% hit rate, changes in performance over time, or comparisons across ages, become more challenging to assess.

Finally, while both the GNG and SST have been used with developmental populations to assess reactive inhibition (i.e., in response to a cue), other tasks utilize a *proactive* inhibition paradigm, known as the Antisaccade Task (AST; e.g., Velanova et al., 2008). Proactive differs from reactive inhibition in that participants actively prepare their task response as part of an internal goal representation instead of simply reacting to stimulus presentation (see Aron, 2011). This kind of inhibitory task may be an attractive alternative for populations that have difficulties in keeping pace with a reactive inhibition task, which generally rely on relatively rapid stimulus delivery to maintain the pre-potent response. Additionally, the AST relies on the same modality for both input and output in the task (i.e., eye movements). As such, this task avoids potential differences in the development or integrity between the visual input and motor output systems that tasks such as the GNG or SST rely on.

2.2. Motivational tasks

While basic cognitive tasks are generally thought to measure "cool" cognitive processes, tasks which involve reward, and often an element

of risk, such as the Balloon Analog Risk Task (BART; Lejuez et al., 2002) or the Monetary Incentive Delay task (MID; Knutson et al., 2001), measure "hot", motivational processes. When examining the development of neural processes involved in reward and risk, a number of considerations should go into task design. One of the most important is selecting a reward that is consistently motivating across all waves of data collection. While the most common approach is to use the same amount of money/points/sugar across all participants, this may be problematic across repeated measures for several reasons. First, it is likely a questionable assumption that a unit of reward ($1, 1 point, 1 piece of candy) has the same subjective value for a participant across different ages (e.g., $1 likely has a greater impact on the finances of an 11- versus a 15- or 20-year-old; see Geier et al., 2009; Galvan, 2010). Secondly, repeated exposure to the same motivator can cause the individual to devalue it at later relative to earlier waves. As such, it could be expected that participants will show reduced motivation for a stable reward over time. For instance, we may see longitudinal declines in reward anticipation (e.g., lower ventral striatum activation) across waves on the MID, where participants must successfully respond to a target presented at a variable interval after a cue in order to obtain rewards (or avoid punishments). However, this reduced anticipation effect may not be driven by developmental changes in reward circuitry, but rather because the threshold of reward that will motivate the participant has changed. Unless this process is specifically the one of interest, a change in reward-value thresholds can present a confound in comparing within-person change because the task is not comparing neural activation to rewards of the same subjective value across development. One way to account for this change is to increase (or decrease as appropriate) the reward value being used to motivate participants in the task. Similar to between-wave changes in GNG difficulty, however; this approach may be limited in accounting for individual differences in subjective reward value. Other research has utilized points instead of concrete rewards (monetary or otherwise) to help account for differences in the subjective value of reward across development (Paulsen et al., 2015; McCormick and Telzer, 2017). However, there unfortunately does not yet exist an equivalent to the adaptive difficulty of the SST for reward valuation which can account for individual differences in subjective value. In lieu of this, subjective value can be assessed by qualitative questions that measure participants' motivation to complete the reward task (e.g., Paulsen et al., 2015) and controlled for in estimating the neural effects of reward/punishment motivation.

Reward tasks often also suffer from similar confounds as the GNG, where the main behavioral metric of inhibitory control also determines the number of events available to estimate the neural signal associated with inhibition. This is especially true when reward motivation is paired with risky behavior. In the BART, participants make increasingly risky decisions (i.e. pumps) in order to earn more points. However, on the neural level, this results in riskier individuals having more data to estimate the neural effects associated with increasing risk and reward. Across repeated measures, this means that the efficacy with which a person's neural signature of reward (or risk) can be estimated fluctuates with their behavioral propensity to pump. While no ideal solution yet exists for this problem of estimation accuracy, researchers should be aware that this may bias their results and the interpretations that can be drawn from them.

2.3. Affective tasks

Although cognitive and motivational tasks have traditionally received a great deal of attention in longitudinal neuroimaging, there has been a relatively recent explosion of interest in the neural development associated with emotional and social processes. Similar to the addition of motivational aspects to a task, the inclusion of affective (and especially face-related) information can present challenges for assessment through longitudinal neuroimaging. Like with rewards, participants can habituate to affect, decreasing their arousal to both positive and negative information. For instance, participants completing an emotional reactivity task using the International Affective Picture System (IAPS) images (Lang et al., 1999) across multiple waves can become used to pictures that were initially affectively arousing. This can result in reduced neural signatures of affective processes, but because of habituation rather than an interesting developmental process. As a solution, the set of pictures a participant views can be changed between waves; however, this can call into question the nature of neural effects seen longitudinally. Neural reactivity may be caused by changes in the types of images viewed or in low-level features that vary between picture sets rather than changes in affective processing *per se*. One solution that can straddle this divide may be to show participants different samples of the same set of images across waves. While the within-person effect may still be biased, this should be controlled for at the group level by randomizing which images are seen at earlier versus later waves for each person.

Another consideration when using faces to generate affective responses is whether to use the same set of faces across all waves of data collection or to use developmentally matched faces at each wave. While using the same faces across all ways has the advantage of controlling for visual and affective features that are almost impossible to match across face datasets, there may be developmentally relevant changes in the ability to recognize or propensity to respond to affect in faces of the same versus different age. In other words, children may show different neural and behavioral responses to adolescent compared to child faces than they will when they are adolescents, and vice versa. If using developmentally appropriate faces at each wave, care should be taken to match faces on a range of features including gender, race, subjective arousal, and low-lying visual components (e.g., luminance, intensity).

2.4. Social decision making tasks

Tasks which are meant to assess complex social processes have unique challenges for longitudinal assessment. Social interaction tasks are many times composite tasks which incorporate aspects of cognitive, motivation, and affective tasks. As such, most of the previous considerations discussed also apply to these tasks. However, some additional concerns present themselves when adding social components. The first is that exposing participants to social situations can carry some increased risk of negatively impacting participant affect or self-esteem. Paradigms that employ deception and experimentally manipulated social rejection, such as Cyberball (Williams and Jarvis, 2006) or the Chatroom Task (Guyer et al., 2009), can cause distress to participants and it is not generally considered ethical to let participants leave the scan session without being debriefed about the deception involved. This presents a significant barrier to longitudinal collection of these tasks, although some groups have done so by making participants aware at the onset that the game is a simulation (i.e. they are not playing with real people). However, this approach has the disadvantage of a reduced ability to measure the process of interest.

Other tasks that are more tractable for longitudinal acquisition, include social decision making tasks that often include a component of deception, which can be a challenge across multiple measurements. Tasks such as the Ultimatum (Sanfey et al., 2003) or Public Goods (Ledyard, 1995) often require the belief that the participant is playing or being watched by one or several other *real* people. Deception paradigms that install this belief range from simple instructions to complex scripts that involve "introductions" between the participants and simulated others. The success of these tasks depends in large part on participants buying into this cover story, particularly at later time points when the participants may be more suspicious due to repeated exposure. Importantly, the more complex a cover story, the greater care will need to be taken to ensure participant buy-in. The impact of suspicion on participants' behavior and neural signatures over time is difficult to quantify; however, there should be some kind of

contingency plan to address participants' questions or suspicions, as well as a recognition that these sorts of tasks may experience various levels of attrition across repeated measures. Care should be taken to only include participants who buy into the deception at all waves, and when debriefing, some metric of how much each participant believed the deception should be included. These steps can help ensure that neural effects are not biased by a decreased believability at later waves.

Of course, one potential solution to this problem is for studies to utilize real individuals, either that are already known to the participants (e.g., parent or friends), or who are introduced and interact directly with the participant during the course of the study (e.g., Diaconescu et al., 2014). This approach not only has the advantage of removing the need for deception, but can also increase the salience of the social processes that underlie a given task. While there are significant challenges in terms of resources and the logistics of more people needing to attend experimental sessions, these approaches can often lend a degree of ecological validity that is difficult to replicate when social tasks involve viewing or interacting with unknown others.

2.5. Other task considerations

No matter the task chosen, participants are only completely naïve to the paradigm and scan environment once. At all subsequent waves, participants come burdened with experience, assumptions, and expectations for a given task. This change in experience can decrease participants' uncertainty about the task parameters. For instance, a participant at later waves may pump more for early balloons in the BART, not because they are inherently riskier, but because they learned that a certain minimum number of pumps are generally safe at an earlier wave. Alternatively, participants can show practice effects on difficult tasks such as the SST through repeated exposure, independent of any developmentally relevant neural maturation. Changes in participants' competence and comfort with a task can affect the signal-to-noise ratio in neural estimation across waves because of reductions in anxiety or uncertainty rather than developmental change. While some amount of this is inevitable, adequate training at early waves can help to ameliorate this effect, especially when testing younger children who may also be more likely to be affected by things outside of the task itself (e.g., scanner anxiety).

Learning effects are not constrained to between waves; they can also occur within a session. When participants complete multiple runs of the same task, or even across blocks within a longer task, the behavioral and neural effects elicited by later trials can be very different than effects elicited by earlier trials. For instance, we have shown how adolescents' behavior and neural responses change across a task session during the BART (e.g., McCormick and Telzer, 2017) or SST (McCormick & Telzer, in press). As such, it is important to design tasks in such a way that performance within a session can be examined. Controlling for within-session change can help to validate between-session changes as developmental in nature instead of experience-related.

One additional consideration that impacts any task choice is the time between measurements within a longitudinal study. While one year follow-ups are the most commonly used in developmental samples, other longitudinal designs may utilize shorter between-measurement intervals. Shorter intervals likely exacerbate challenges stemming from learning and habituation, as participants are more likely to retain explicit memories about the task environment and any strategies they might have used if applicable; however, they may offer advantages for other processes. For instance, changes in the subjective reward value of a reinforcer is more likely to be comparable when individuals are measured across weeks or months compared with a year or more. While the interval between sessions is likely determined by a number of external factors (e.g., funding, participant and/or researcher burden), it should nevertheless be consistent with theoretical expectations about the kinds of changes that could be observed across waves.

While we have focused, thus far, on considerations related to repeated exposure to task environments, one method that has gained increasing popularity in recent years is task-free, resting state scans. Resting state offers some unique advantages compared with task-based scans, since it avoids many of the problems related to difficulty, habituation, or learning effects we have discussed previously. In addition, resting-state networks are thought to reflect the long-term accumulation of life experiences on functional connectivity that are not constrained by any particular task context. However, this does not mean that resting state does not pose its own challenges. Because of the task-free environment, it is more difficult to attribute specific functional significance to changes over time in resting-state networks. Furthermore, in terms of measurement, it is likely that there are developmental differences in the experience of resting-state that may lead to differences in the data measured (see Rosenberg et al., 1997; Raschle et al., 2009 for examples). Resting state is particularly susceptible to scanner anxiety and movement, in part because there is no active task context to maintain attention and focus. As these factors are often associated with age, ensuring equitable levels of comfort and movement across measurements is key for resting-state analyses. For the same reasons, resting state is also more likely to be impacted by other state level changes (e.g., sleep, mood, previous scans). Alternative protocols to resting state data have been proposed that may alleviate some of these issues (e.g., Inscapes; Vanderwal et al., 2015), which researchers might consider, especially if working with high-movement populations.

In summary, the selection of an appropriate task is crucial for the success of a longitudinal neuroimaging study. While the advantages and disadvantages of all the tasks available to choose from could fill this entire issue, the above considerations should have broad applicability to a variety of commonly used and validated paradigms beyond those described in this section. Selecting a particular task inherently comes with trade-offs, yet we hope that this discussion can help to elucidate some of the most common and important factors to consider when making that decision.

2.6. Considerations for reliability

Although there may be age-related changes in cognitive, affective or social processes, there is also some level of stability within individuals. Understanding the reliability (i.e., stability) of task-related neural effects is essential for truly examining developmental changes. Without carefully considering the reliability of neural patterns, changes may be attributed to development when, in fact, the change is due to subject-related, task-related, or scanner-related variance (see Herting et al. this issue). Beyond task considerations that may affect reliability across waves due to, for example, learning, motivational changes, or data processing choices (e.g., Shirer et al., 2015), subject-related variance can greatly impact the reliability of neural signal across time. Subject-related variables may include hormonal rhythms, sleep, and stress, variables that change developmentally. Researchers have utilized short-term, within-subject designs to examine how differences in these subject-related variables impact neural signal. For instance, utilizing daily experience sampling methods to assess stress, Galvan and colleagues (e.g., Uy and Galvan, 2017) brought adolescents into the lab on a low-stress day and a high-stress day, scans which occurred within about a week of each other. Adolescent boys showed damped PFC activation during a gambling reward task on high stress days. This is an important finding given that adolescence is a time of heightened stress (Dahl and Gunnar, 2009). If one does not collect and control for stress, it is possible that developmental differences in an adolescent sample are due to changes in stress and not to development, per se. Others have shown within-person effects as a function of time of day in adults, such that neural activation differs depending on the time of the scan (e.g., Bryne et al., in press). Thus, when collecting longitudinal fMRI data, these sources of variance should be standardized and controlled for, both between and within-subjects, to improve test-retest reliability of results

and ensure that differences found across waves are due to development and not to sources of variance that affect reliability. See Herting (et al., current issue) for an excellent review on the current state of test-retest reliability in developmental samples.

3. Modeling longitudinal change in neuroimaging studies

Given the growth of longitudinal imaging datasets, a timely question is how to handle this data in a way that assumptions are met and valid inferences can be drawn. Group analysis of functional neuroimaging datasets typically follow a two-step approach. The first-level analyses center on the individual level, whereas the second-level analyses center on the group level, in which all effects of interest are summarized and tested across subject. Whereas first-level modeling in longitudinal imaging may come with some challenges, such as registration, the second-level step brings particular statistical challenges for a longitudinal neuroimaging design. That is, the standard General Linear Model (GLM) is appropriate for designs where there is only one scan per subject; the basic type of statistical tests implemented in the main software packages are not well-suited for longitudinal data. When pursuing a whole-brain analysis on longitudinal data, analyses that account for repeated measures are required, which are implemented, to some degree, in all major software packages for imaging data (e.g., FSL: http://fsl.fmrib.ox.ac.uk/fsl/; SPM: http://www.fil.ion.ucl.ac.uk/spm/; AFNI: https://afni.nimh.nih.gov). In an attempt to highlight a few of the most often used approaches we discuss group-level longitudinal data-analysis tools in well-known software packages, as well as the popular mixed-effect modeling approach. Note that we do not attempt to provide a full statistical overview of the packages (there are a number of excellent papers which do that, e.g. Chen et al., 2013, 2014; McFarquhar et al., 2016), but instead aim to provide a concise overview of these approaches.

3.1. FMRIB's FSL

In FSL, longitudinal analyses typically include a regressor per subject in addition to group- or condition-level differences. This set-up is explained with clear examples on the FSL wiki (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/GLM), together with their assumptions and caveats. The main constraints and assumptions with this method center on completeness and sphericity of the data. That is, this GLM, as implemented in FSL, cannot handle missing data and thus requires each subject to have complete data at all time points. Also, a repeated-measures analysis in Feat (FSL's toolbox for functional data-analysis) assumes sphericity, and more specifically a compound-symmetric structure in each voxel. That is, both the variance across time points and the covariance between all time points is assumed to be constant. The assumption of compound symmetry is relatively strict, but probably reasonable when the data is sampled in a regular way. Yet, for more complex (e.g., more than 2 time points or irregularly sampled) longitudinal designs such an assumption may be problematic. With this whole-brain approach, researchers can explore questions such as which brain regions show a significant change over time (as in a paired *t*-test) and/or which voxels respond to a significant group x time interaction. Given that subject-specific regressors are included, however, researchers cannot examine direct questions of between-subject effects. That is, group-level differences, such as a main effect of group, can only be interrogated with an additional analysis that averages the measures within-subject. Moreover, due to using a univariate GLM approach, *t*-test of within-subject effects that do not add up to 0, such as a main effect of one condition, cannot be validly examined. Also, the inclusion of multiple within-subject factors is problematic in the GLM framework, because of the complexity in variance partitioning. Particularly, handling multiple within-subject factors leads to an incorrect differentiation in error partitioning, and consequently flawed results. See for a more detailed discussion on this topic Chen et al. (2014).

3.2. SPM

In SPM (http://www.fil.ion.ucl.ac.uk/spm/), testing simple repeated measures effects in a second-level analysis can be done with the 'flexible factorial' ANOVA module. This module allows the creation of a design matrix that, similar to Feat, includes a *subject* variable, leading to the inclusion of a regressor per subject on top of other within-subject (e.g., time) and group variables (e.g., children/adults; patients/control). As in FSL, flexible factorial models can include only subjects that have complete data at all time points. In contrast to FSL, SPM has a method for correcting violations of sphericity (Glaser and Friston, 2007), although the estimated covariance structure is assumed to be the same for each voxel. That these assumptions hold in simple longitudinal designs seems well established, yet whether this is also the case in more complex, multi-time point, data is a point of ongoing debate. A flexible factorial ANOVA allows researchers to probe questions of which brain regions show a main effect over time, or a time x group interaction. Again, similar to FSL, between-subject effects, such as a main effect of group, can be inquired only with additional analyses, *t*-test of within-subject effects that do not add up to 0, cannot be validly examined, and multiple within-subject factors are hard to investigate because of flaws in error partitioning (e.g., Chen et al., 2014).

In sum, both FSL and SPM software packages allow for testing whole-brain effects in longitudinal design, yet do not allow for investigating the between-subject effects, and are limited for more complex extensions (multiple within-subject levels or multiple between-subject factors). Finally, they have limitations in dealing with various cases of missing data in longitudinal studies.

3.3. GLM Flex and GLM Flex Fast2

GLM Flex Fast2 is the most recent version of GLM Flex and these MATLAB packages are often the next step if SPM's flexible factorial model is not sufficient. These packages use a GLM framework, but unlike SPM's flexible factorial model, they can accommodate multiple between-subject variables in addition to within-subject factors, and models can include up to 6 factors. Examples of potential designs are included on the package website (http://mrtools.mgh.harvard.edu/). Another advantage of using these packages are that covariates can be included, although they are limited to between-subject variables. To estimate the variance-covariance structure, GLM Flex uses the same procedure as SPM and has the option to correct for sphericity violations. In contrast to FSL and SPM, GLM Flex packages partition the error terms rather than pooling them and may therefore be less susceptible to inflated false positive rates than repeated-measures analyses implemented in these programs (McLaren et al., 2011). However, like FSL and SPM, GLM Flex packages are limited in that they cannot handle missing data or include within-subject covariates, such as age or pubertal status. Further, because models with more than one within-subject factor have not yet been validated within this framework, it is unclear whether these models are correctly implemented in GLM Flex packages.

3.4. AFNI

The primary advantage of using AFNI for longitudinal modeling is that it includes packages for repeated-measures that do not rely on a univariate GLM framework and are therefore substantially more flexible. 3dMVM (Chen et al., 2014) is the primary tool for ANOVA-style analyses and utilizes a multivariate GLM framework. This modeling approach separates within- and between-subject factors, allowing for an unlimited number of factors. In addition to multiple within- and between-factors, any number of continuous between-subject variables can be included, making 3dMVM amenable to most GLM, ANOVA or ANCOVA designs. However, while groups can be unequal, data from all subjects must be complete. To verify the assumption of sphericity in 3dMVM, the variance-covariance structure is estimated from the data

and Mauchly's test is conducted to assess whether this assumption has been violated. If sphericity is violated, voxel statistics can be adjusted via the sphericity correction option. For more complex designs that include within-subject covariates or have missing data, linear mixed-effects modeling can be conducted using 3dLME (Chen et al., 2013), which is further discussed in the next paragraph.

3.5. Mixed effect modeling

Multilevel analyses can be considered an extension of repeated-measures designs and have the advantage that both the starting point (intercept) and slope (change over time) for each individual are taken into account. Multilevel techniques are commonly available in analysis programs such as R (R Core Team, 2016), specifically in packages *lme4* (Bates et al., 2015) and *nlme* (Pinheiro et al., 2016), but also in MPlus, STATA, SAS and SPSS. Growth curve multilevel models, also known as mixed models, mixed-effect models, or hierarchical linear models, can estimate the average growth trajectory for the whole sample (i.e., group-level or fixed effects) on a measure of interest (e.g., brain activation), and model the variation in these effects (i.e., random effects). The latter may provide valuable information on within-subject change and the heterogeneity of, for instance, brain activation. This multilevel growth curve modeling technique has the flexibility to model data that have been collected at uneven intervals, and does not require all participants to have the same number of data points, thereby flexibly accounting for missing data (although it is assumed that cases are missing 'at random'; Little, 1988; Curran et al., 2010). Developmental changes in brain function have been tested with growth curve models by using a polynomial approach. That is, longitudinal studies have compared linear (Age), quadratic (Age^2) and/or cubic (Age^3) effects of Age to best describe a regional pattern of brain activation, but inverse, logarithmic or exponential functions can also be tested (Curran et al., 2010). Effects of multiple time-variant covariates (e.g., testosterone levels, learning performance) and/or time-invariant covariates (e.g., sex) are easily included.

A whole brain approach of a mixed effect model is AFNI's 3dLME, which is built around the nlme and lme4 mixed effects modeling packages (Bates et al., 2015; Pinheiro et al., 2017) in R (R Core Team, 2016). 3dLME is a flexible package that can accommodate most designs. Any number of categorical or continuous variables can be included as fixed-effects, and random effects (e.g., subject intercepts and/or slopes) are specified directly, although variance-covariance structures cannot be customized. Due to the fact that within-subject covariates can be included, this modeling approach is ideal for longitudinal studies seeking to model development (e.g., chronological age) as a continuous rather than categorical variable. Further, because any number of continuous within-subject variables can be included in the model, developmental trends (e.g., linear, quadratic, cubic trajectories) and their interactions with other variables can be assessed (assuming there are a sufficient number of observations per subject to warrant these analyses). Another advantage is that models allow the inclusion of subjects with missing data, provided that these adhere to the missing-at-random (MAR) assumption. (see Matta, Flournoy, & Byrne, this issue for important considerations regarding missing data). In addition to traditional experimental designs, 3dLME can also be used to assess the stability or reliability of effects by calculating intraclass correlations across the whole-brain (Herting et al., this issue).

An alternative to whole-brain analyses are ROI-based methods that target a predetermined neuroanatomical structure or functional region. These methods are appropriate for more targeted hypotheses, where the main questions of interest are what pattern of change in activity occurs in a particular brain region or what variables explain the change in brain activation over time. Although this method is spatially coarser, it is flexible in its use. That is, data are reduced to a single dependent value for a region of interest (or multiple if several brain regions are examined). This data reduction allows for a range of statistical analyses, among which multilevel techniques are popular. The advantage of an ROI approach is that they allow for the most flexible designs, including any type of model that can be specified in R, including not only SEM models, but also models for questions that require modeling change as a latent variable or to examine brain changes as a mediator or moderator of associations.

4. Review of recent developmental longitudinal fMRI studies

The field of longitudinal neuroimaging in developmental samples is still in its infancy. To date, longitudinal fMRI studies have generally focused on the following three topics: 1) investigating developmental trajectories and 2) predicting (current or future) behavior based on brain measures. In this section, we first review research that has investigated developmental trajectories, focusing on studies using accelerated longitudinal designs with an ROI approach. We then discuss studies focusing on longitudinal data analyses with the goal to predict future behavioral outcomes. Finally, we conclude with a detailed review of three studies. We have selected a key paper from each of our research groups and go into detail about the decisions made in terms of task considerations, sampling, and analysis. This is not meant to be an exhaustive review but instead to highlight key considerations from several recent papers. See Table 1 for summary of studies included in the review.

4.1. Developmental trajectories

Several longitudinal fMRI studies have compared developmental trajectories using an accelerated longitudinal design. An accelerated longitudinal design may include cohorts of several ages that are followed with a certain time lag, or more of an ad-hoc sampling strategy including multiple measurements of a continuous age range. One of the first fMRI studies using an accelerated longitudinal design (Ordaz et al., 2013; N = 123, 1–6 measurements, age 9–26 years) used an inhibitory anti-saccade task to compare group-level linear, quadratic and loglinear trajectories on a set of pre-specified ROIs. Results suggested that motor-related regions developed relatively early (as these showed no age-related changes), executive control regions showed a decrease in activation into early adulthood, whereas error-related activity increased into early adulthood in the anterior cingulate cortex. Another study (Paulsen et al., 2015; N = 82 subjects with 2 (N = 49) or 3 (N = 33) measurements, age 10–20 years) compared group-level linear and quadratic trajectories during inhibition and reported linear developmental trajectories in cortical regions (including VLPFC and frontal eye fields) and a quadratic effect (adolescent dip) in bilateral posterior parietal cortex. A study using a similar approach but focusing on reward processing (Braams et al., 2015; N = 299, 8–27 years) compared linear, quadratic and cubic trajectories on the group-level and found that striatum activity for rewards (modeled at feedback onset) showed a quadratic trajectory, peaking in adolescence.

Rather than using predetermined developmental trajectories (e.g. linear, quadratic), another possibility is to investigate the best fitting function regardless of its shape (spline function), which allows for a more flexible description of developmental change. This method was implemented by Simmonds et al. (2017; N = 57, 1–9 measurements, age 8–30), who used mixed-effects spline growth models. The authors examined the longitudinal development of neural activity for working memory in ROIs based on a separate cross-sectional sample, which has several advantages such as including task-relevant regions without biasing towards stability or change. Results indicated that regional activation changed most profoundly up to mid-adolescence, increasing in visual regions during encoding and retrieval, and decreasing in frontal and subcortical regions during working memory maintenance, which was related to task performance such as precision error and decision latency. A potential consideration when using spline models instead of polynomial functions (e.g., linear, quadratic, and cubic

Table 1
Longitudinal fMRI studies discussed in this review.

	N	T	Included scans per subject per T (1/2/3 +)	Ages over Ts	Design	Regions	Statistics	Modality	Time between Ts	Dataset
Functional brain data										
Ordaz et al. (2013)	139	6	123/79/58/29/12/1	9–26	Accelerated	ROIs	HLM/LMM	Antisaccade	Variable	Pittsburgh Luna
Paulsen et al. (2015)	187	3	82/49/33	10–20	Accelerated	ROIs	LMM	Antisaccade		
Simmonds et al. (2017)	129	9	57/57/57/41/25/12/6/3/1	10–30	Accelerated	WholeBrain/ROIs	LMM	Memory guided Antisaccade		
Braams et al. (2015)	299	2	249/238	10–20	Accelerated	ROIs	LMM	Reward task	2yrs	Braintime Crone
Braams and Crone (2017)	299	2	249/238	8–27	Accelerated	ROIs	LMM	Social reward task		
Peters et al. (2016)[a]	**299**	**2**	**208/208**	**10–20**	Accelerated	**ROIs**	**LMM**	**Feedback learning**		
Peters et al. (2017)	299	2	274/231	8–27	Accelerated	ROIs	LMM	Resting state		
Koolschijn et al. (2011)	32	2	32/32	8–27	Accelerated	WholeBrain	SPM	Feedback learning	3.5yrs	Leiden Crone
Qu et al. (2015a)[a]	**24**	**2**	**22/22**	**15–18**	Longitudinal	**WholeBrain**	**SPM/GLMflex**	**BART**	**1.5yrs**	**UCLA Telzer**
Qu et al. (2015b)	24	2	23/23	15–18	Longitudinal	WholeBrain	SPM	BART		
Qu et al. (2016)	24	2	23/23	15–18	Longitudinal	WholeBrain	SPM	BART		
McCormick et al. (2016)	23	2	20/20	14–15	Longitudinal	WholeBrain	SPM/GLMflex	GoNogo	1yr	UNC Telzer
McCormick et al. (2017)	23	2	20/20	14–15	Longitudinal	WholeBrain	SPM/GLMflex	GoNogo		
Gabard-Durnam et al. (2016)	23	2	23/23	7–15	Accelerated	ROIs	AFNI	Emotional faces and Resting state	2yrs	UCLA Tottenham
Pfeifer et al. (2013)[a]	**27**	**2**	**27/27**	**10–13**	Longitudinal	**WholeBrain**	**SPM**	**Self-evaluation**	**3yrs**	Pfeifer/Dapretto
Uy and Galván (2017)	45	2	45/45	15–17/25–30	Longitudinal	WholeBrain	FSL	Risk taking	~2wks	UCLA Galván
Ullman et al. (2014)	62	2	62/62	6–20	Accelerated	WholeBrain/Multimodal	SVR	Visual-Spatial Working Memory	2yrs	Karolinska Klingberg

N = total number of subjects, T = time point, ROI = regions of interest, LMM = linear mixed models, HLM = hierarchical linear model, SVR = support vector regression.
[a] Discussed in detail in the section Detailed Reviews and Critiques.

trajectories) is that although they represent the data best, they are sometimes more difficult to link to theories of development and may suffer from overfitting the data.

Together, these studies provide some of the first evidence demonstrating longitudinal changes in neural activation across wide age ranges, showing that developmental trajectories depending on the cognitive function (i.e., inhibition versus reward processing) and brain region. The studies in this section furthermore show that ROI analyses are important for comparing different shapes of developmental trajectories and there is currently no whole-brain alternative widely available. Assessing developmental trajectories is important even when researchers are not primarily interested in the shape of a developmental trajectory (but focus for instance on brain-behavior correlations and control for age), because there could be changes with age that are non-linear in nature. On the other hand, studies covering a wide age-range typically use an accelerated longitudinal design. These designs are well-suited to test effects on the group level (i.e., fixed effects), describing within-subject change is more ambiguous, because the level of change should be interpreted vis-a-vis age at starting point. Although some steps have already been made for assessing individual's developmental change using longitudinal research, a lot more research is needed in order to be able to confirm or disconfirm theoretical perspectives on development, and to examine whether findings obtained from cross-sectional and group-level comparisons also reflect within-person changes over time (e.g., McCormick et al., 2017).

4.2. Brain as predictor of future behavior

Aside from assessing developmental trajectories, researchers have also used longitudinal fMRI data to investigate the relation between neural measures and behavioral measures. In these studies, brain-behavior relations and not necessarily age-related changes are the main interest. This method of "brain as predictor of behavior" (e.g., Berkman and Falk, 2013) or "neuroforecasting" (e.g., Genevsky et al., 2017) allows researchers to test how neural processes inform later behavior. Importantly, developmental research contributes to this method in that longitudinal data can make predictions on brain-behavior relations stronger by testing whether changes in fMRI activity over time also correspond with changes in behavior over time. In some studies, both brain and behavior measures are assessed at two or more time points, but sometimes initial brain measures are used to predict future behavioral outcomes and vice versa.

4.3. Changes in brain as mediator of longitudinal behavioral change

Several studies have used mediation approaches to confirm that brain-related changes mediate a longitudinal behavioral relation. Qu and colleagues (Qu et al., 2015b) tested adolescents (N = 23, age 15–17) who completed the Balloon-Analog-Risk task (BART) at two time points. Whole-brain analyses with changes in parent-child interactions as a regressor showed that greater increases in positive parent-child interactions were related to larger decreases in striatum and DLPFC activity to reward over time. Mediation analyses showed that changes in ventral striatum activity (based on a functional ROI) mediated the relation between changes in positive parent-child relationships and changes in risk-taking behavior. Another study in the same sample showed that changes in ventral striatum activity mediated the relation between T1 parental depression and changes in risk-taking behavior over time (Qu et al., 2016). Finally, a study (N = 22 adolescents, age 14) found that longitudinal change in VLPFC activation during a GNG task mediated the relation between negative family relations at T1 and changes in risk taking over time (McCormick et al., 2016).

4.4. Developmental sequences

Another interesting longitudinal approach for brain-behavior

relations is to investigate developmental sequences. Although truly causal experiments can often not be performed in humans, longitudinal studies can shed light on the order of two processes: which precedes what? For instance, it has often been reported that there are neural differences between adolescents who consume high amounts of alcohol and adolescents who do not. However, it is difficult to disentangle whether alcohol is the cause of these neural differences, or whether these differences were already present prior to initiation of alcohol use (i.e., having a "risky brain"). As an example of this method, Peters et al. (2017) examined adolescents (N = 274 at T1 and N = 231 at T2, age 12–27) and found that reduced amygdala-OFC resting state connectivity predicted increased alcohol use two years later, but alcohol use did not predict later amygdala-OFC connectivity, suggesting that reduced amygdala-OFC connectivity precedes alcohol use and may bias adolescents towards risky behavior. In future studies, cross-lagged panel models could be an excellent method to test questions of this nature (see Meeus, 2016 for a review).

4.5. Cross-modal prediction

Longitudinal studies can also be used to determine how neural measures from different modalities influence each other. For instance, one of the long-standing questions in functional MRI research is how resting state fMRI networks develop into mature and stable networks. Gabard-Durnam et al. (2016) hypothesized that recurrent task-based neural activity shapes the development of resting-state networks (the long-term phasic molding hypothesis). They employed an accelerated longitudinal design in adolescents (N = 23 for prospective subsample, age 4–18) and tested the prospective associations between task-elicited amygdala activity and resting state connectivity of the amygdala two years later. Age-effects for both task-based and resting state amygdala connectivity were found in an overlapping region in medial PFC. Moreover, stimulus-elicited amygdala-mPFC connectivity at T1 unidirectionally predicted resting-state amygdala-mPFC connectivity. This suggests that stimulus-elicited connectivity precedes resting state connectivity and may be crucial in shaping resting state networks during development. Similar methods could be employed to test the relation between brain structure and function.

4.6. Machine learning approaches

Finally, when using neural measures to predict future behavioral outcomes, it is becoming increasingly recognized that machine learning approaches and cross-validation are crucial (Gabrieli et al., 2015). Most prior studies assessed variation in a neural measure and used that to predict variation in a behavioral measure. If we aim to identify practical implications of neuroscience (such as to discover neuromarkers that predict behavioral outcomes), it is crucial to assess how well a prediction model will perform on a new individual (unseen data). These approaches can be used for both binomial outcomes (e.g. developing depression or not) and continuous outcomes (e.g. future mathematics performance). Machine-learning can be performed using many different algorithms. Often-used algorithms include support vector machines and random forest algorithms. Briefly, a support-vector machine algorithm works by finding where a line (decision plane) should be placed to make the best separation between two groups of data. Random forest algorithms work by combining multiple 'decision trees', which each make decisions on how two groups of data can be distinguished from each other. Given that machine-learning algorithms cannot handle temporal dependency well, they are typically not suited for tracking within-subject change and using the full value of longitudinal designs (see, however, Aksman et al., 2016 for a longitudinal application), but can answer questions of prediction from cross-sectional data.

For instance, using a prediction approach, a recent study showed that decreased activity in the striatum and DLPFC ROIs to anticipated rewards predicted whether novelty-seeking adolescents (N = 144, selected from the IMAGEN sample, MRI at one time point) would later develop problematic drug use (Büchel et al., 2017). The researchers used a support vector machine to confirm that out-of-sample prediction accuracy was higher for a model including brain measures compared to a model with only behavioral measures. In another study, rather than ROI-based analyses, Ullman et al. (2014) used a voxel-wise support-vector machine model to test which regions in the brain could predict future working memory capacity two years later. Participants (N = 62, 6–20 years, MRI at one-time point) performed a visuospatial working memory task in the MRI scanner. The whole-brain results indicated that current working memory capacity was related to frontoparietal activity, but interestingly, future working memory capacity could be predicted from the structure and activity in the basal ganglia and thalamus.

4.7. Detailed reviews and critiques

Next, we provide detailed reviews of three longitudinal neuroimaging studies from each of our groups. We selected these studies in order to unpack the many different types of decisions that are made in terms of task design, behavioral analyses, and neuroimaging analyses. By reviewing our own work, we are able to be self-critical and relatively comprehensive in detailing the pros and cons of the diverse analyses and quality checks that we conducted.

In a recent study (Peters et al., 2016), we used data from an accelerated longitudinal design in which 299 participants between ages 8–25 at time point 1 (T1) were tested twice within a 2-year interval ("Braintime" sample). Only participants with full data for a feedback-learning task at two time points were included (N = 208, age 8–27). The main aim was to investigate the developmental trajectory of neural activity in frontoparietal regions associated with cognitive control.

The choice to only include participants with sufficient quality data at both time points was made to have a truly longitudinal sample. The Braintime sample had a high retention rate (only 13 of 299 participants did not participate at T2), but more were excluded for the neuroimaging part at T2 because of braces (N = 32). With hindsight, we could have performed our growth trajectory analyses while including participants with missing data at one of the two waves. It is important to carefully consider the balance between including only full participant data while at the same time avoiding data loss. We did not use data imputation methods but this is often done in behavioral developmental studies (for a review see Enders, 2013).

Another methodological consideration is that the task had to be understandable for young children (8 years) but also not too easy for the adults. We succeeded in designing a task that was appropriate for young children, but also still showed age-related improvements. Towards adulthood developmental changes were limited, possibly because this was simply a task with a clear developmental endpoint (as is the case for many cognitive processes) or, alternatively, it could reflect a ceiling effect. A possible solution is using performance-adaptive paradigms to ensure that age-related changes are not influenced by performance differences, but these have the inherent downside that the development of behavioral performance cannot be investigated, and may therefore be less ecologically valid.

For our fMRI analyses, we focused on targeted ROIs rather than whole-brain analyses. Because of the accelerated longitudinal design y, testing change from T1 to T2 is likely to be age-dependent and a direct whole-brain comparison between T1 and T2 may mask age-related change (e.g. if older participants show no changes from T1 to T2, this may mask changes in children from T1 to T2). Therefore, whole-brain comparisons may only be informative in accelerated designs when a difference score between time points is tested against a baseline age. Direct whole-brain comparisons between two time points can on the other hand be suitable for cohort designs.

We performed mixed-model analyses using the R package nlme. We focused on four anatomically defined ROIs associated with feedback learning: DLPFC, superior parietal cortex (SPC), supplementary motor

area (SMA) and ACC. Anatomical ROIs are not biased towards one time point (such as when using ROIs based on T1) or biased towards stability (ROIs based on a T1 and T2 average). A disadvantage of anatomical ROIs is that they may not always be task-relevant, although in our paper these anatomical regions showed high overlap with whole-brain task activity. Other options would have been to include ROIs based on coordinates from meta-analyses, prior independent studies, or programs such as Neurosynth. It could be argued that it is necessary to use multiple-comparisons for multiple ROIs. A good option might be Bonferroni correction adjusted for the correlation between the ROIs (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm) (Perneger, 1998; Sankoh et al., 1997).

Based on dual-systems models we tested whether cognitive brain regions show group-level linear changes with development, or whether there were adolescent-specific (quadratic) or adolescent-emergent (cubic) effects (Casey, 2015). Our results revealed that activation in DLPFC and SPC was best characterized by a quadratic trajectory peaking/leveling off towards late adolescence/early adulthood, whereas SMA showed a linear increase and ACC showed a linear decrease with age. One complicating factor is that the tails of the distribution may have a disproportionately large influence on the best model. In an accelerated longitudinal study both the youngest and oldest ages will include less data-points than the remaining dataset. Adding confidence intervals to predicted trajectories is therefore important, and caution is warranted when interpreting effects for which plateaus appear within the confidence interval. Additionally, the absolute pattern of change will depend on the included age-range of the sample, and observed 'peaks' may depend on the included age-range. Data from a third time point will allow a more dense sampling of young adult participants and may further confirm the group-level quadratic trajectories in this dataset.

We also showed that reliability for neural activity in DLPFC, SPC, SMA and ACC was fair to good (ICC-value with absolute agreement, average measure > 0.4; Cicchetti, 2001). Although we could not test for differences in ICC between age groups, there was no clear age-related pattern of reliability (e.g., lower reliability in the youngest age group). ICC-values, anecdotally, were markedly higher in this study compared to both subcortical and cortical activity during social and affective processes in the same sample (Braams and Crone, 2017; Braams et al., 2015). Lower reliability for affective processes or tasks compared to cognitive processes or tasks would be interesting to confirm in future studies.

An additional aim was to investigate whether changes from T1 to T2 were explained not only by age, but also by individual differences in task performance, working memory capacity and cortical thickness. We performed mixed-model stepwise regression analyses with neural activity as dependent variable and the best-fitting age model as the first predictor, and tested whether performance, working memory, and cortical thickness explained additional variance in neural activity over age alone. The results indicated that task performance explained variance in DLPFC and SPC activity, and cortical thickness explained variance in SMA activity. We added age as a first step in the regression because all variables were correlated with age. However, it is conceptually also essential to consider whether by controlling for age we are not removing variance of interest. It is important to theorize about what age effects actually entail, as increased cognitive performance and structural maturation are both highly intertwined with development. In this way age-differences in functional maturation may be driven by cognitive performance and structural maturation, which could be tested in a longitudinal mediation framework.

In a recent study (Qu et al., 2015a), we used a cohort design to examine change in neural activation during risk taking across two time points. This study utilized a small (N = 22) sample of 15–17 year olds at the first time point, who were then scanned again approximately 1.5 years later. Adolescents completed the BART task, during which they made decisions to inflate a virtual balloon. Pumps at each time point were associated with earning 25 cents. A benefit of focusing on one age range is that we did not have to worry that the subjective value of 25 cents varied across the developmental sample, which is often the case when utilizing a large age range that includes children and adults. Nonetheless, by including only 2 times points within a small age range, this study is unable to examine developmental trajectories, and we cannot determine whether our effects reflect adolescent-specific phenomena, as any changes we see could reflect change regardless of age (e.g., adults may evidence similar changes across the 1.5 year period).

The first decision to make is how to model the task chosen. We describe here the behavioral analyses, and below will go into detail about the fMRI analyses. We used adjusted pumps, which represents the average pumps on balloons that did not result in an explosion. The reason for excluding pumps on balloons that exploded is that such balloons can artificially constrain participants' risk level, particularly when balloons explode early. Adjusted pumps is one of the most common ways to analyze behavior on the task, and was done by the original creators of the BART (Lejuez et al., 2002), but others have used number of explosions (e.g., Braams et al., 2015), total points earned (e.g., Peper et al., 2013), or parameters estimated at a trial-by-trial level to represent a learning index (e.g., McCormick and Telzer, 2017). Flexibility in modeling the task allows for creative analyses, but also can limit reproducibility and increase researchers' degrees of freedom (see Harden et al., 2017).

It is essential in longitudinal fMRI research, particularly when using tasks for which learning or habituation are not the key construct, to ensure that changes in performance over time are not due to learning on the task. For instance, participants may pump more at later waves, not because they are riskier, but because they learned the task constraints at earlier waves and know balloons may explode at a certain level. To test this possibility, we examined change in behavior within one time point by dividing the task into halves (i.e., do participants change their behavior within a task session) as well as change in behavior across time (i.e., do participants change their behavior over the longitudinal period. We found no group-level change in performance in terms of average adjusted pumps within or across sessions. This provides some evidence that learning does drive our main effects. However, several studies have found important developmental changes in learning on the BART. For instance, in separate cross-sectional samples, we found age-related linear increases in learning across the first and second thirds of the task within one session in 4–26 year olds (Humphreys et al., 2016), as well as age-related linear increases in learning on a trial-by-trial level in 8–17 year olds (McCormick and Telzer, 2017). Therefore, learning on the BART certainly occurs, and age may be an important predictor of such learning, which has significant implications for longitudinal research, particularly accelerated longitudinal designs with a wide age range.

Another important consideration is to ensure that the task taps the construct of interest. The BART is designed to measure risk taking, and prior research has shown that more pumps on the task are associated with real-life risk taking (e.g., Lejuez et al., 2002). Yet, some have argued that the association between task performance and self-reports is negligible or may even be measuring different constructs (see Harden et al., 2017). To test the ecological validity of the task, and to ensure that change in behavior over time was meaningful, we correlated adolescents' risk-taking behavior on the BART with their self-reported risk taking. Importantly, adolescents who showed increases in risk taking on the BART also reported increases in risk taking in their daily lives. This quality check provides some confidence that the BART is tapping into risk taking, and that individual differences in developmental changes we see in task performance and corresponding neural signal are relevant for adolescents' real life risk taking.

The second decision is how to model the task at the neural level, which includes both fixed effects first-level models and random effects group-level models. Consistent with the behavioral analyses, we modeled the adjusted pumps, excluding pumps that resulted in an explosion

as well as separately modeling the outcome (cash-out or explosion). One problem with the task is that the number of events within a condition depends on behavior, so those who are more risky have more data to model. This can be particularly problematic in longitudinal analyses, as the estimates at one time point may be more reliable due to having more data. Therefore, one may find significant changes over time that are due to noise rather than to real developmental changes. To control for this, we included a parametric modulator on each trial in the subject's first-level models, representing the number of pumps for each balloon. We modeled the parametric modulator as a control variable rather than a variable of interest. Thus, when computing difference scores across the two time points, neural activation represents differences in the recruitment of a brain region, controlling for the level of risk within each trial at each time point. When examining developmental change, or utilizing a large age range, including the parametric modulator at the trial level may be the most appropriate, but it can be limiting as well if the assumed functional form of the parametric modulator is mis-specified (e.g., a linear modulator but a quadratic neural response).

At the group-level, we computed difference scores to examine neural activation across the two time points. Briefly, we found that activation in the VLPFC decreased across the two time points in the whole sample (i.e., "normative" developmental change), and adolescents who showed decreases in the VLPFC and VS over time as well as decreases in MPFC-VS coupling showed declines in risk taking (i.e., individual differences in developmental change). While these findings are important in identifying changes in the functional role of the VLPFC and MPFC in adolescent risk taking, the results are limited. In particular, with only two time points and one age range, difference scores were calculated, which do not tell us about the starting point of development (i.e., someone who changes from 10 to 13 units appears the same as someone who changes from 0 to 3 units). The starting point (or intercept) can be key in identifying developmental processes. For instance, in another study, we utilized a similar analytic approach for a two-wave examination of cognitive control-related neural development (McCormick et al., 2017). All adolescents were 14 years old at the first time point and were scanned a second time one year later. Even though they were all the same age, adolescents' starting point of VLPFC activation varied across the sample. Importantly, adolescents' VLPFC activation at wave 1 was important for determining how their VLPFC changed across the year − those who started high in VLPFC activation tended to show declines over the year, whereas those who started low tended to show increases − and this change predicted their risky behavior. These data highlight the importance of understanding the intercept or initial level of the brain unit in order to better capture developmental changes.

In addition to ensuring the behavioral effects are not due to learning or repetition, we ran parallel analyses at the whole-brain level to test for learning. First, we examined change in neural activation within the first time point by dividing the task into halves. Consistent with the behavioral analyses, there was no change in neural activation within the session, even at a liberal statistical threshold. Next, we identified regions which showed statistically significant change over time, which included the VLPFC. We created a functional ROI of this region and extracted activation from within the first session during the first and second halves of the task and computed a difference score. We then reran our whole brain, longitudinal analyses including this variable as a covariate. While this method may not be the optimal approach, it tests whether longitudinal changes in the VLPFC hold when controlling for within session changes that could be driven by learning or habituation.

In another relatively recent study (Pfeifer et al., 2013), a combination of the decisions and features mentioned in the two studies previously reviewed in depth can be observed. This study reports results from a cohort of N = 27 10-year-olds (M age = 10.1, SD = 0.35) who provided high-quality fMRI data at both the initial and a subsequent time point three years later (M age = 13.1, SD = 0.33). These early adolescents completed a block-design task in which they were asked to evaluate the extent to which short trait phrases representing the social and academic domains described themselves, or a familiar fictional other (Harry Potter).

In this study, we chose to include only participants that could provide a 'complete' dataset, so those who did not have high quality data at wave 1 or wave 2 were excluded (whether due to motion, attrition, lack of compliance with task, lack of exposure to the other social target, computer malfunction, random assignment to alternative fMRI tasks, etc.). Because this represents a combination of random and non-random reasons for missingness, we should anticipate this would bias our estimates of brain activity elicited by the task and change therein over time to some degree (for an illustration of the extent of bias in this sample, see Matta et al., this issue). However, this choice was made because at the time we were unaware of any alternative approach that would allow us to conduct a whole-brain search with an imbalanced dataset. In retrospect it would be ideal to provide comprehensive descriptions of these reasons for exclusion, as well as model using alternative approaches that allow the use of all available data instead of the subset which is 'complete.'

Another consideration was how to model the task. The phrases were blocked by target (self or other) and domain (social or academic), and in each block positive and negative phrases were pseudorandomly distributed along with some null events. This presented the option of modeling the design by events, or in a mixed fashion with both sustained (block) and transient (event) effects. We chose to model blocks only, for several reasons. First, while the phrases were clearly positive or negative, whether each event produced a positive or negative evaluation of the target was idiosyncratic to participant responses (in other words, agreeing that a positive phrase described oneself or disagreeing that a negative phrase described oneself both would constitute positive self-evaluations). Participants also demonstrated a strong positivity bias, so there were far more positive than negative evaluations of targets, for both the self and other social target. For most participants, this would have resulted in too few events of some types to model successfully at the first level in an event-related manner. However, in subsequent studies we modeled similar paradigms as events rather than blocks, collapsing across whether the evaluation produced by combining the valence of the phrase and the participants' response was ultimately positive or negative (e.g., Jankowski et al., 2014). The two modeling approaches seemed to provide similar patterns of activity in cortical midline structures, although it is difficult to compare precisely because the other social target used varied across the two studies.

We identified main effects of target (self versus other) in cortical midline structures and ventral striatum using an F test, but also confirmed the involvement of these regions across both waves using a conjunction analysis. This analytical step was valuable since it demonstrated that at both timepoints, activity elicited in those regions was significant (rather than being driven primarily by one or the other wave). We also examined changes over time in the contrast between self and other by testing the statistical interaction between time (wave 1 versus 2) and target (self versus other) at the whole-brain level. This produced activity in ventromedial PFC. We then interrogated this specific cluster to determine whether the changes were more pronounced for the social or academic domain, as well as whether the changes were related to pubertal development, an analysis that was made feasible by the tight age range from which we sampled. An alternative would have been to define some ROIs implicated in self-evaluation, whether anatomically or functionally using a meta-analytical approach or relying on prior independent clusters. However, cortical midline structures are typically not well defined anatomically, and since the functionally independent approaches would be almost completely defined by adult samples we opted to interrogate the specific cluster that evidenced change within our sample.

5. Future directions and conclusion

Longitudinal studies are quickly rising in the developmental cognitive neuroscience field, notwithstanding the difficulties in designing, acquiring, and testing longitudinal change. In this review, we first highlighted the most common and important factors of longitudinal task design, such as learning, ceiling effects, and task-reliability. Although most of our reviewed factors are important for all studies, they may be even more prominent, and some even unique, to longitudinal imaging. Then, we discussed group-level statistical models as available, including both whole-brain and region-of-interest approaches. It is imminent that this field is changing fast and additions in existing statistical packages as discussed here will provide researchers with more flexibility to analyze their longitudinal imaging data. For instance, a new R package neuropointillist (http://ibic.github.io/neuropointillist/; see Madhyastha et al., this issue) has recently been launched that allows one to use flexible estimates in longitudinal models of brain function, such as latent growth curve models and mixture models, on a whole brain level at a multicore machine. Another example is by Kievit et al. (this issue) who advocate for structural equation models for longitudinal imaging data and have made analysis code available in freely available software package R such as Lavaan (Rosseel, 2012). Such exciting new initiatives will allow researchers to better test specific questions regarding group-level and within-person developmental change. Moving forward, the field of developmental neuroscience will need to grapple with the fact that individuals may not follow group-level trajectories of development, as well as apply statistical techniques that appropriately disambiguate between- and within-subject level effects (e.g., Curran and Bauer, 2011; Preacher et al., 2010). As the field progresses, new longitudinal initiatives in large consortia are becoming available in Europe in developmental samples, such as generation R), the Consortium on Individual Development (CID; www.individualdevelopment.nl), and in the United States, such as the Pediatric Longitudinal Imaging, Neurocognition and Genetics (PLING; www.chd.ucsd.edu/research/pling.html), and the recently launched Adolescent Brain and Cognitive Development study (ABCD; http://addictionresearch.nih.gov/abcd-study). Use of these larger longitudinal cohort studies will allow researchers to boost sample sizes across the full age range from childhood through adolescence and adulthood. Up to now, most longitudinal imaging studies have used only two time points most with only one to two years between measurements. Although these studies allow for estimates of within-individual change, they are limited in their ability to test individual's trajectory of change for which more time points will be necessary (Singer and Willet 2003; Curran et al., 2010). Moreover, many of these datasets to come may help to disambiguate within and between-subject change which is most clear-cut when individuals of one starting age are followed throughout development. Lastly, a particularly exciting new initiative is the possibility for replication across longitudinal samples, which has started for structural brain development between childhood and adulthood across multiple separate longitudinal samples (Tamnes et al., 2017), but has not yet been done in functional MRI studies.

In conclusion, our goal was to highlight the current state of the developmental cognitive neuroscience field focusing on methodological considerations for longitudinal task-based fMRI. And whereas the field of structural longitudinal imaging is surging, there are limited developmental longitudinal studies that are taking a functional network approach, nor are there multimodal studies combining longitudinal information of structure, function, and behavior across development. It seems that current tools are not designed to test change in network level connectivity, which leaves an important gap to fill. In sum, by designing studies with the most sophisticated tools and using the most appropriate statistical analyses, longitudinal neuroimaging can get to the most important question for developmental scientists: how and why do people differ in the way they develop?

Acknowledgements

EHT receives support from the Department of Psychology and Neuroscience at UNC-CH, with a grant from the National Science Foundation (SES-1459719) and National Institutes of Health (R01-DA039923). ACKD receives support from an Open Research Area (ORA) grant (ASTA 464-15-176) and a Leiden Aspasia grant. JHP receives support from the National Institutes of Health (R01-MH107418).

References

Aksman, L.M., Lythgoe, D.J., William, S.C.R., Jokisch, M., Monninghoff, C., Streffer, J., et al., 2016. Making use of longitudinal information in pattern recognition. Hum. Brain Mapp. 37, 4385–4404.
Aron, A.R., 2011. From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. Biol. Psychiatry 69 (12), e55–e68.
Büchel, C., Peters, J., Banaschewski, T., Bokde, A.L., Bromberg, U., Conrod, P.J., et al., 2017. Blunted ventral striatal responses to anticipated rewards foreshadow problematic drug use in novelty-seeking adolescents. Nat. Commun. 8.
Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., 2015. Lme4: Linear Mixed-effects Models Using Eigen and S4, 2014. R Package Version 1. pp. 4.
Berkman, E.T., Falk, E.B., 2013. Beyond brain mapping: using neural measures to predict real-world outcomes. Curr. Directions Psychol. Sci. 22 (1), 45–50.
Braams, B.R., Crone, E.A., 2017. Longitudinal changes in social brain development: processing outcomes for friend and self. Child Dev. 88 (6), 1952–1965.
Braams, B.R., van Duijvenvoorde, A.C., Peper, J.S., Crone, E.A., 2015. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. J. Neurosci. 35 (18), 7226–7238.
Casey, B.J., 2015. Beyond simple models of self-control to circuit-based accounts of adolescent behavior. Annu. Rev. Psychol. 66, 295–319.
Chen, G., Saad, Z.S., Britton, J.C., Pine, D.S., Cox, R.W., 2013. Linear mixed-effects modeling approach to FMRI group analysis. Neuroimage 73, 176–190.
Chen, G., Adleman, N.E., Saad, Z.S., Leibenluft, E., Cox, R.W., 2014. Applications of multivariate modeling to neuroimaging group analysis: a comprehensive alternative to univariate general linear model. Neuroimage 99, 571–588.
Cicchetti, D.V., 2001. Methodological commentary the precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. J. Clin. Exp. Neuropsychol. 23 (5), 695–700.
Crone, E.A., Elzinga, B.M., 2015. Changing brains: how longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. Wiley Interdiscip. Rev.: Cogn. Sci. 6 (1), 53–63.
Curran, P.J., Obeidat, K., Losardo, D., 2010. Twelve frequently asked questions about growth curve modeling. J. Cogn. Dev. 11 (2), 121–136.
Dahl, R.E., Gunnar, M.R., 2009. Heightened stress responsiveness and emotional reactivity during pubertal maturation: implications for psychopathology. Dev. Psychopathol. 21 (01), 1–6.
Diaconescu, A.O., Mathys, C., Weber, L.A., Daunizeau, J., Kasper, L., Lomakina, E.I., et al., 2014. Inferring on the intentions of others by hierarchical Bayesian learning. PLoS Comput. Biol. 10 (9), e1003810.
Durston, S., Thomas, K.M., Yang, Y., Uluğ, A.M., Zimmerman, R.D., Casey, B.J., 2002. A neural basis for the development of inhibitory control. Dev. Sci. 5 (4), F9–F16.
Enders, C.K., 2013. Dealing with missing data in developmental research. Child Dev. Perspect. 7 (1), 27–31.
Gabard-Durnam, L.J., Gee, D.G., Goff, B., Flannery, J., Telzer, E., Humphreys, K.L., et al., 2016. Stimulus-elicited connectivity influences resting-state connectivity years later in human development: a prospective study. J. Neurosci. 36 (17), 4771–4784.
Gabrieli, J.D., Ghosh, S.S., Whitfield-Gabrieli, S., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. Neuron 85 (1), 11–26.
Galvan, A., 2010. Adolescent development of the reward system. Front. Hum. Neurosci. 4.
Geier, C.F., Terwilliger, R., Teslovich, T., Velanova, K., Luna, B., 2009. Immaturities in reward processing and its influence on inhibitory control in adolescence. Cereb. Cortex 20 (7), 1613–1629.
Genevsky, A., Yoon, C., Knutson, B., 2017. When brain beats behavior: neuroforecasting crowdfunding outcomes. J. Neurosci. 37 (36), 8625–8634.
Glaser, D., Friston, K., 2007. Covariance components. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier, pp. 140–147.
Guyer, A.E., McClure-Tone, E.B., Shiffrin, N.D., Pine, D.S., Nelson, E.E., 2009. Probing the neural correlates of anticipated peer evaluation in adolescence. Child Dev. 80 (4), 1000–1015.
Harden, K.P., Kretsch, N., Mann, F.D., Herzhoff, K., Tackett, J.L., Steinberg, L., Tucker-Drob, E.M., 2017. Beyond dual systems: a genetically-informed: latent factor model of behavioral and self-report measures related to adolescent risk-taking. Dev. Cogn. Neurosci. 25, 221–234.
Herting, M.M., Gautam, P., Chen, Z., Mezher, A., Vetter, N.C., 2017. Test-retest reliability of longitudinal task-based fMRI—implications for developmental studies. Dev. Cogn. Neurosci. http://dx.doi.org/10.1016/j.dcn.2017.07.001.
Humphreys, K.L., Telzer, E.H., Flannery, J., Goff, B., Gabard-Durnam, L., Gee, D.G., et al., 2016. Risky decision making from childhood through adulthood: contributions of learning and sensitivity to negative feedback. Emotion 16 (1), 101.
Jankowski, K.F., Moore, W.E., Merchant, J.S., Kahn, L.E., Pfeifer, J.H., 2014. But do you

think I'm cool?: Developmental differences in striatal recruitment during direct and reflected social self-evaluations. Dev. Cogn. Neurosci. 8, 40–54. http://dx.doi.org/10.1016/j.dcn.2014.01.003.
Kievit, R.A., Brandmaier, A.M., Ziegler, G., van Harmelen, A.L., de Mooij, S.M., Moutoussis, M., ... Lindenberger, U., 2017. Developmental cognitive neuroscience using Latent Change Score models: a tutorial and applications. Dev. Cogn. Neurosci In press, corrected proof.
Knutson, B., Adams, C.M., Fong, G.W., Hommer, D., 2001. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J. Neurosci. 21 (16), RC159.
Koolschijn, P.C.M., Schel, M.A., de Rooij, M., Rombouts, S.A., Crone, E.A., 2011. A three-year longitudinal functional magnetic resonance imaging study of performance monitoring and test-retest reliability from childhood to early adulthood. J. Neurosci. 31 (11), 4204–4212.
Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1999. International Affective Picture System (IAPS): Technical Manual and Affective Ratings. The Center for Research in Psychophysiology, University of Florida, Gainesville, FL.
Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Roth, A.E., Kagel, J. (Eds.), Handbook of Experimental Economics. Princeton University Press, Princeton, pp. 111–194.
Lejuez, C.W., Read, J.P., Kahler, C.W., Richards, J.B., Ramsey, S.E., Stuart, G.L., et al., 2002. Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). J. Exp. Psychol.: Appl. 8 (2), 75.
Li, C.S.R., Huang, C., Constable, R.T., Sinha, R., 2006. Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. J. Neurosci. 26 (1), 186–192.
Little, R.J., 1988. A test of missing completely at random for multivariate data with missing values. J. Am. Stat. Assoc. 83 (404), 1198–1202.
Louis, T.A., Robins, J., Dockery, D.W., Spiro, A., Ware, J.H., 1986. Explaining discrepancies between longitudinal and cross-sectional models. J. Chronic Dis. 39 (10), 831–839.
Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., ... McLaughlin, K.A., 2017. Current methods and limitations for longitudinal fMRI analysis across development. Dev. Cogn. Neurosci In press, corrected proof.
Matta, T., Flournoy, J.C., Byrne, M., 2017. Making an unknown a known unknown: missing data in longitudinal neuroimaging studies. Dev. Cogn. Neurosci. http://dx.doi.org/10.1016/j.dcn.2017.10.001.
McCormick, E.M., Telzer, E.H., 2017. Adaptive adolescent flexibility: neurodevelopment of decision-making and learning in a risky context. J. Cogn. Neurosci. 29 (3), 413–423.
McCormick, E.M., Qu, Y., Telzer, E.H., 2016. Adolescent neurodevelopment of cognitive control and risk-taking in negative family contexts. Neuroimage 124, 989–996.
McCormick, E.M., Qu, Y., Telzer, E.H., 2017. Activation in context: differential conclusions drawn from cross-sectional and longitudinal analyses of adolescents' cognitive control-related neural activity. Front. Hum. Neurosci. 11.
McFarquhar, M., McKie, S., Emsley, R., Suckling, J., Elliott, R., Williams, S., 2016. Multivariate and repeated measures (MRM): a new toolbox for dependent and multimodal group-level neuroimaging data. Neuroimage 132, 373–389.
McLaren, D.G., Schultz, A.P., Locascio, J.J., Sperling, R.A., Atri, A., 2011. Repeated-measures designs overestimate between-subject effects in fMRI packages using one error term. In 17th Annual Meeting of Organization for Human Brain Mapping 26–30.
Meeus, W., 2016. Adolescent psychosocial development: a review of longitudinal models and research. Dev. Psychol. 52 (12), 1969.
Menon, V., Adleman, N.E., White, C.D., Glover, G.H., Reiss, A.L., 2001. Error-related brain activation during a Go/NoGo response inhibition task. Hum. Brain Mapp. 12 (3), 131–143.
Ordaz, S.J., Foran, W., Velanova, K., Luna, B., 2013. Longitudinal growth curves of brain function underlying inhibitory control through adolescence. J. Neurosci. 33 (46), 18109–18124.
Paulsen, D.J., Hallquist, M.N., Geier, C.F., Luna, B., 2015. Effects of incentives, age, and behavior on brain activation during inhibitory control: a longitudinal fMRI study. Dev. Cogn. Neurosci. 11, 105–115.
Peper, J.S., Koolschijn, P.C.M., Crone, E.A., 2013. Development of risk taking: contributions from adolescent testosterone and the orbito-frontal cortex. J. Cogn. Neurosci. 25 (12), 2141–2150.
Perneger, T.V., 1998. What's wrong with bonferroniadjustments. Br. Med. J. 316, 1236–1238.
Peters, S., Van Duijvenvoorde, A.C., Koolschijn, P.C.M., Crone, E.A., 2016. Longitudinal development of frontoparietal activity during feedback learning: contributions of age, performance, working memory and cortical thickness. Dev. Cogn. Neurosci. 19, 211–222.
Peters, S., Peper, J.S., Van Duijvenvoorde, A.C., Braams, B.R., Crone, E.A., 2017. Amygdala-orbitofrontal connectivity predicts alcohol use two years later: a longitudinal neuroimaging study on alcohol use in adolescence. Dev. Sci. 20 (4).
Pfeifer, J.H., Kahn, L.E., Merchant, J.S., Peake, S.J., Veroude, K., Masten, C.L., et al., 2013. Longitudinal change in the neural bases of adolescent social self-evaluations: effects of age and pubertal development. J. Neurosci. 33 (17), 7415–7419. http://dx.doi.org/10.1523/JNEUROSCI.4074-12.2013.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., 2016. R Core Team (2016) Nlme: Linear and Nonlinear Mixed Effects Models. R Package Version 3. pp. 1–128. Available at https.cran.r-project.org/web/packages/nlme/index.html (Accessed 7 July).
Preacher, K.J., Zyphur, M.J., Zhang, Z., 2010. A general multilevel SEM framework for assessing multilevel mediation. Psychol. Methods 15 (3), 209.
Qu, Y., Galvan, A., Fuligni, A.J., Lieberman, M.D., Telzer, E.H., 2015a. Longitudinal changes in prefrontal cortex activation underlie declines in adolescent risk taking. J. Neurosci. 35 (32), 11308–11314.
Qu, Y., Fuligni, A.J., Galvan, A., Telzer, E.H., 2015b. Buffering effect of positive parent-child relationships on adolescent risk taking: a longitudinal neuroimaging investigation. Dev. Cogn. Neurosci. 15, 26–34.
Qu, Y., Fuligni, A.J., Galván, A., Lieberman, M.D., Telzer, E.H., 2016. Links between parental depression and longitudinal changes in youths' neural sensitivity to rewards. Soc. Cogn. Affect. Neurosci. 11 (8), 1262–1271.
R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/.
Raschle, N.M., Lee, M., Buechler, R., Christodoulou, J.A., Chang, M., Vakil, M., et al., 2009. Making MR imaging child's play-pediatric neuroimaging protocol, guidelines and procedure. J. Visual. Exp. 29.
Rosenberg, D.R., Sweeney, J.A., Gillen, J.S., Kim, J., Varanelli, M.J., O'hearn, K.M., et al., 1997. Magnetic resonance imaging of children without sedation: preparation with simulation. J. Am. Acad. Child Adolesc. Psychiatry 36 (6), 853–859.
Rosseel, Y., 2012. Lavaan: An R Package for Structural Equation Modeling and More. Version 0. Ghent University, Ghent, Belgium, pp. 5–12 (BETA).
Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., Cohen, J.D., 2003. The neural basis of economic decision-making in the ultimatum game. Science 300 (5626), 1755–1758.
Sankoh, A.J., Huque, M.F., Dubey, S.D., 1997. Some comments on frequently used multiple endpoint adjustment methods in clinical trials. Stat. Med. 16, 2529–2542.
Shirer, W.R., Jiang, H., Price, C.M., Ng, B., Greicius, M.D., 2015. Optimization of rs-fMRI pre-processing for enhanced signal-noise separation, test-retest reliability, and group discrimination. NeuroImage 117, 67–79.
Simmonds, D.J., Hallquist, M.N., Luna, B., 2017. Protracted development of executive and mnemonic brain systems underlying working memory in adolescence: a longitudinal fMRI study. Neuroimage 157, 695–704.
Singer, J.D., Willet, J.B., 2003. A framework for investigating change over time. Appl. Longitud. Data Anal.: Model. Change Event Occur. 3–15.
Tamnes, C.K., Herting, M.M., Goddings, A.L., Meuwese, R., Blakemore, S.J., Dahl, R.E., et al., 2017. Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. J. Neurosci. 37 (12), 3402–3412.
Ullman, H., Almeida, R., Klingberg, T., 2014. Structural maturation and brain activity predict future working memory capacity during childhood development. J. Neurosci. 34 (5), 1592–1598.
Uy, J.P., Galvan, A., 2017. Sleep duration moderates the association between insula activation and risky decisions under stress in adolescents and adults. Neuropsychologia 95, 119–129.
Vanderwal, T., Kelly, C., Eilbott, J., Mayes, L.C., Castellanos, F.X., 2015. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. Neuroimage 122, 222–232.
Velanova, K., Wheeler, M.E., Luna, B., 2008. Maturational changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. Cereb. Cortex 18 (11), 2505–2522.
Williams, K.D., Jarvis, B., 2006. Cyberball: a program for use in research on interpersonal ostracism and acceptance. Behav. Res. Methods 38 (1), 174–180.